A new method for estimating the parameter

Romain François Peltier, Jacques Lévy Véhel

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

A new method for estimating the parameter of fractional Brownian motion

Romain François PELTIER - Jacques LEVY VEHEL

N° 2396
Novembre 1994

PROGRAMME 4

Rapport de recherche

A new method for estimating the parameter of fractional Brownian motion

Romain François PELTIER, Jacques LEVY VEHEL

Programme 4 — Robotique, image et vision
Action Fractales

Rapport de recherche n° 2396 — Novembre 1994 — 27 pages

Abstract: We introduce a new estimator of the parameter of a fractional Brownian motion and prove that it is strongly consistent. We also give rates of convergence, asymptotic confidence intervals and tests of fit for this estimator. Simulation experiments show that our methods perform well with respect to other techniques available in the literature.

Key-words: fractional Brownian motion, fractal, continuous process.

Romain François Peltier: Université de PARIS VI - INRIA: e-mail: peltier@bora.inria.fr
Jacques Lévy Véhel: e-mail: jlv@bora.inria.fr

Unité de recherche INRIA Rocquencourt
Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
Téléphone : (33 1) 39 63 55 11 – Télécopie : (33 1) 39 63 53 30

INRIA ROCQUENCOURT

Une nouvelle méthode d'estimation du paramètre du mouvement Brownien fractionnaire

Résumé : Nous introduisons un nouvel estimateur du paramètre du mouvement Brownien fractionnaire et montrons qu'il est fortement consistant. De plus, nous donnons les vitesses de convergence, des intervalles de confiance asymptotiques et proposons un test d'adéquation à un mouvement Brownien fractionnaire associé à notre estimateur. Les simulations montrent que notre méthode est meilleure que celles existantes dans la littérature.

Mots-clé : mouvement Brownien fractionnaire, fractale, processus continu.

1. Introduction

The fractional Brownian motion of index H $(0 < H < 1)$ was defined by Mandelbrot and Van Ness (1968) as the stochastic integral, for $t \geq 0$

$$X_H(t) = \frac{1}{\Gamma(H+1/2)} \left\{ \int_{-\infty}^{0} [(t-s)^{H-1/2} - (-s)^{H-1/2}] dW(s) + \int_{0}^{t} (t-s)^{H-1/2} dW(s) \right\},$$

where $\{W(s), -\infty < s < \infty\}$ denotes a Wiener process extended to the real line.
It may be shown (see e.g. Mandelbrot (1968) p. 425) that the definition above gives, up to an equality in probability, *the only Gaussian process* $\{X(t),\, t \geq 0\}$ *such that* $X(0) = 0$ *and there exists* $K > 0$ *such that for any* $t \geq 0$ *and* $h > 0$, $X(t+h) - X(t)$ *follows a normal* $N(0, K^{2H} h^{2H})$ *distribution* [K^H is usually called its *scale factor*].

This process was first introduced to model time series describing the motions of Nil's river flow. Subsequently, its practical applications were extended into many domains, such as quantum mechanics, economics, medical image analysis and many more. It turns out that its statistical study has so far focused on the estimation of the parameter H, neglecting the important problem of fitting the model.
A survey of this topic is to be found in Beran (1992) and also Flandrin *et al* (1991) who compare various estimators of H: the most widely used method for estimating H given the observation of a sampled version $\{X_H(\frac{kT}{n}), 0 \leq k \leq n\}$ of $\{X_H(t), 0 \leq t \leq T\}$ is the Hurst analysis (see e.g. Hurst (1955)). However, its performances are far from being optimal and several other estimators have been proposed for H since the introduction of this method. The estimator introduced by Fox and Taqqu (1986) is based on a maximum likelihood analysis and has an asymptotically optimal rate of consistency, namely $O(1/\sqrt{n})$ when $n \to \infty$. However its pratical implementation raises some technical difficulties involving for example the computation of inverses of large covariance matrices. Some other estimators are based on the fractal dimension of the process trajectory. In these cases, the estimates are obtained via slope measurements in log-log plot of length versus scale. Flandrin *et al* (1991) do not establish the theoretical rate of consistency, but give empirical evidence based on simulated data that some of the estimates they consider and in particular the Higuchi (1988) method perform well.
In this paper, we propose new estimators for H, based on absolute moments of the increments as the one of Higuchi (1988). We evaluate their efficiency with respect to the estimators previously considered in the literature and show that they have a rate of convergence equal to $O(n^{-1/2}(\log n)^{-1})$ when $n \to \infty$ for $H \in (0, 3/4)$. Finally, we describe a test of fit based on our estimators, and show some experimental evidence that our estimator generally behaves better than other ones.
The remainder of our paper is organised as follows. In Section 2, we recall some basic facts about *box dimension*, and introduce an estimator for this quantity. In Section 3, we introduce two families of estimators for H and prove their almost sure consistency. In Section 4, we describe asymptotic confidence intervals for H and propose a test of fit to fractional Brownian motion. In Section 5, we present some simulation results showing that our methods perform well in practice and that the computation of our estimators is simple and fast.

2. Box dimension

To describe our estimators, we need to introduce the following notation. For any $\delta > 0$, we consider the set of *δ-mesh squares* in $\mathbb{R}^2$ of the form $[i\ \delta, (i+1)\ \delta] \times [j\ \delta, (j+1)\ \delta]$ with i, j integers. For any bounded subset F of $\mathbb{R}^2$, we denote by $N_\delta(F)$ the number of δ-mesh squares which intersect F. The box dimension of F is then defined by (see e.g. Falconer (1990) p.41)

$$\dim(F) = \lim_{\delta \to 0} \left(\frac{\log N_\delta(F)}{-\log \delta} \right), \tag{2.0}$$

whenever this limit exists.

The box dimension of the graph of X_H is characterized in Lemma 2.1 below (see e.g. Falconer (1990) p.246):

Lemma 2.1 *With probability 1 a fractional Brownian motion X_H with index H has graph $\{(t, X_H(t)) : t \in [0,1]\}$ with box dimension equal to $2 - H$.*

In the sequel, we will use fractal dimension as a synonym for box dimension. If $X : [0,1] \to \mathbb{R}$ is a continuous function, then it is well known that there exists a sequence $X_n : [0,1] \to \mathbb{R}$, $n = 1, 2, \ldots$ of polygonal functions which converges uniformly to X on $[0,1]$. Moreover, we may and do assume that the vertices of the graph of each X_n are of the form $\{(\frac{k}{n}, X_n(\frac{k}{n}))$, $0 \le k \le n\}$, with $X_n(0) = 0$. Setting $\mathcal{X}_{k,n} = X_n\left(\frac{k}{n}\right)$ for $0 \le k \le n$, we will set for convenience $\mathcal{X}_{(n)} = (\mathcal{X}_{1,n}, \ldots, \mathcal{X}_{n,n})$.

We will make an instrumental use of the following lemma in the derivation of our estimators. Given a polygonal function X_n as above, we set for each $n \ge 2$,

$$L_n = \sum_{i=1}^{n-1} |\mathcal{X}_{i+1,n} - \mathcal{X}_{i,n}|. \tag{2.1}$$

We will say that a continuous function $X : [0,1] \to \mathbb{R}$ has box dimension D if its graph $\{(t, X(t)), t \in [0,1]\}$ is a subset of $\mathbb{R}^2$ whose box dimension is D.

Lemma 2.2 *Let $X : [0,1] \to \mathbb{R}$ be a continuous function with box dimension D, and let $(X_n)_{n \ge 1}$ be a sequence of polygonal functions as above which converges uniformly to X on $[0,1]$. The following results hold*

(1) *If X is constant, then $D = 1$;*

(2) *If X is non constant then*

$$D = 1 + \lim_{n \to \infty} \frac{\log L_n}{\log(n-1)}. \tag{2.2}$$

Proof. By definition of the box dimension we easily obtain the first statement of the lemma. For the second statement, set $\delta_n = \frac{1}{n-1}$ for each $n \ge 2$, and denote respectively by $N_{\delta_n,n}$ and N_{δ_n} the number of δ_n-mesh squares that intersect the graphs of X_n and X. Let L_n be as in (2.1). We have the obvious inequalities

$$L_n \le N_{\delta_n,n}\ \delta_n \le L_n + (n-1)\delta_n.$$

If X is non constant then there exists an $a > 0$ such that $L_n \geq a$ for all n large enough. Thus

$$1 + \frac{\log L_n}{\log(n-1)} \leq -\frac{\log N_{\delta_n,n}}{\log \delta_n} \leq 1 + \frac{\log(1+L_n)}{\log(n-1)},$$

and

$$\lim_{n\to\infty}\left(-\frac{\log N_{\delta_n,n}}{\log \delta_n}\right) = \lim_{n\to\infty}\left(1 + \frac{\log L_n}{\log(n-1)}\right) = \lim_{n\to\infty}\left(1 + \frac{\log(1+L_n)}{\log(n-1)}\right) \tag{2.3}$$

whenever the last limit exists. Since X_n converges uniformly to X on $[0,1]$ as $n \to \infty$, we have

$$\lim_{n\to\infty}\left(\frac{\log N_{\delta_n,n}}{\log N_{\delta_n}}\right) = 1. \tag{2.4}$$

On the other hand, by the definition (2.0) of the box dimension D of X, we have

$$\lim_{n\to\infty}\left(\frac{\log N_{\delta_n}}{\log \delta_n}\right) = -D. \tag{2.5}$$

A joint application of (2.4) and (2.5) shows that

$$\lim_{n\to\infty}\left(-\frac{\log N_{\delta_n,n}}{\log \delta_n}\right) = D, \tag{2.6}$$

which, when combined with (2.3) yields (2.2). □

3. Estimators of H

In the sequel we will let $K > 0$ be a fixed real and introduce estimators of H based on the observation of a sampled version of the process $\{K^H X_H(t) : 0 \leq t \leq T\}$. For the sake of notational simplicity, we will set throughout $T = 1$ and

$$X_{i,n} = K^H X_H\left(\frac{i}{n}\right) \text{ for } i = 0, 1, ..., n. \tag{3.0}$$

The main results of this section are stated in Proposition 3.2 (where we describe the behavior of a new estimate of the parameter of fractional Brownian motion), Theorem 3.7 (which highlights a property of fractional Brownian motion) and Proposition 3.8 (where we propose another new class of estimates of H).

Recall the well known easy fact (see e.g. Papoulis (1991) p.110).

Fact. *Let Z denote a normal $N(0,1)$ random variable. Then for any $r > 0$*

$$E[|Z|^r] = \frac{2^{r/2}\Gamma\left(\frac{r+1}{2}\right)}{\Gamma(\frac{1}{2})}. \tag{3.1}$$

The following technical proposition will be useful in the sequel. Fix any $0 < H < 1$ and $K > 0$. For each $n \geq 2$, denote by $\mathbb{X}_{1,n}, \dots, \mathbb{X}_{n,n}$ an array of random variables such that, for each $i = 1, \dots, n-1$, $\mathbb{X}_{i+1,n} - \mathbb{X}_{i,n}$ follows an $N(0, {\sigma_n}^2)$ distribution with ${\sigma_n}^2 = \left(\frac{K}{n-1}\right)^{2H}$. For each integer $k \geq 1$, set further

$$\mathbb{S}_n(k) = \frac{1}{n-1}\sum_{i=1}^{n-1} |\mathbb{X}_{i+1,n} - \mathbb{X}_{i,n}|^k.$$

Using (3.1), we have that

$$E_n(k) := E(\mathbb{S}_n(k)) = \frac{2^{k/2}K^{Hk}}{(n-1)^{Hk}} \times \frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{\pi}}.$$

Proposition 3.1 *Under the above assumptions, for each $k \geq 1$ we have*

$$\lim_{n\to\infty} \mathbb{S}_n(k) = 0 \qquad a.s., \tag{3.2}$$

and, for each $0 \leq H' < H$

$$\lim_{n\to\infty}\left((n-1)^{H'} \sup_{1\leq i\leq n-1}(|\mathbb{X}_{i+1,n} - \mathbb{X}_{i,n}|)\right) = 0 \qquad a.s. \tag{3.3}$$

Proof. - In the first place, we recall that if $f(x)$ is a convex fonction of $x \in \mathbb{R}$ then, for any $(x_1, \dots, x_n) \in \mathbb{R}^n$, we have the inequality, $f\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) \leq \frac{1}{n}\sum_{i=1}^{n} f(x_i)$.[This follows from Jensen's inequality $f(E(U)) \leq E(f(U))$ when applied to a random variable U uniformly distributed on $\{x_1, \dots, x_n\}$.] By choosing $f(x) = |x|^p$, we infer from this inequality that, for each $p \geq 1$,

$$|\mathbb{S}_n(k)|^p \leq \frac{1}{n-1}\sum_{j=1}^{n-1} |\mathbb{X}_{j+1,n} - \mathbb{X}_{j,n}|^{kp}.$$

Therefore, for any $\varepsilon > 0$ and $p \geq 1$, the Markov inequality entails that

$$\begin{aligned} P[\mathbb{S}_n(k) > \varepsilon] &\leq \frac{E((\mathbb{S}_n(k))^p)}{\varepsilon^p} \\ &\leq \frac{1}{\varepsilon^p}E\left(\frac{1}{n-1}\sum_{j=1}^{n-1}|\mathbb{X}_{j+1,n} - \mathbb{X}_{j,n}|^{kp}\right) = \frac{1}{\varepsilon^p}E(|\mathbb{X}_{2,n} - \mathbb{X}_{1,n}|^{kp}). \end{aligned} \tag{3.4}$$

Set for convenience

$$c_k = \frac{2^{kp/2}K^{Hkp}\Gamma\left(\frac{kp+1}{2}\right)}{\sqrt{\pi}\varepsilon^p}.$$

Inequality (3.4) in combination with (3.1) entails that

$$P[\mathbb{S}_n(k) > \varepsilon] \leq c_k \frac{1}{(n-1)^{Hkp}},$$

and we conclude by using the Borel-Cantelli lemma with $p > \frac{1}{H}$ to obtain (3.2).

- In the second place, for the proof of (3.3) we start by observing that, for any $\varepsilon > 0$, $0 < H' < H < 1$, $p \geq 1$

$$\begin{aligned} P[(n-1)^{H'} \sup_{1\leq i\leq n-1}(|\mathbb{X}_{i+1,n} - \mathbb{X}_{i,n}|) > \varepsilon] &\leq \sum_{i=1}^{n-1} P[(n-1)^{H'}|\mathbb{X}_{i+1,n} - \mathbb{X}_{i,n}| > \varepsilon] \\ &\leq \frac{(n-1)}{\varepsilon^p}E((n-1)^{H'p}|\mathbb{X}_{i+1,n} - \mathbb{X}_{i,n}|^p). \end{aligned}$$

Set for convenience

$$c'_p = \frac{2^{p/2}K^{Hp}\Gamma\left(\frac{p+1}{2}\right)}{\sqrt{\pi}\varepsilon^p}.$$

We see that

$$P[(n-1)^{H'} \sup_{1\leq i\leq n-1}(|\mathbb{X}_{i+1,n} - \mathbb{X}_{i,n}|) > \varepsilon] \leq c'_p \frac{1}{(n-1)^{(H-H')p-1}},$$

so that, by taking $p > \frac{2}{H-H'}$, (3.3) follows readily from the Borel-Cantelli lemma. □

Recall the notation (3.0) for the sampled process $\{X_{i,n} = K^H X_H\left(\frac{i}{n}\right), 0 \leq i \leq n\}$, and let $k \geq 1$ and $K > 0$ be fixed. We start by the study of the estimators of H given by

$$H_{n,K}(k) = -\frac{\log\left[\sqrt{\pi}S_n(k)/(2^{k/2}\Gamma\left(\frac{k+1}{2}\right))\right]}{k\ \log((n-1)/K)}, \tag{3.5}$$

where

$$S_n(k) = \frac{1}{n-1}\sum_{i=1}^{n-1}|X_{i+1,n} - X_{i,n}|^k. \tag{3.6}$$

Our first result establishes the consistency of $H_{n,K}(k)$ to H.

Proposition 3.2 For each integer $k \geq 1$ and $K > 0$, we have

$$\lim_{n\to\infty} H_{n,K}(k) = H \qquad a.s. \tag{3.7}$$

Proof. In our proof, we will limit ourselves to $K = 1$ and set $H_n(k) = H_{n,1}(k)$, for the sake of notational simplicity. It will become obvious later on that our arguments may be used to cover the case of an arbitrary $K > 0$ at the expense of elementary technicalities. The proof of (3.7) will be achieved via the following five steps.

Step 1. We infer from Jensen's inequality that

$$(S_n(1))^k \leq S_n(k).$$

Step 2. Application of Lemmas 2.1 and 2.2 to $X = X_H$ yields:

$$\lim_{n\to\infty}\left(-\frac{\log(S_n(1))}{\log(n-1)}\right) = H \qquad \text{a.s.}$$

Step 3. In this step, we show that for any $0 \leq H' < H$,

$$\lim_{n\to\infty}\frac{S_n(k)}{(n-1)^{-H'k}} = 0 \qquad \text{a.s.}$$

This statement turns out to be a consequence of (3.2) in Proposition 3.1 together with the straightforward inequalities

$$\inf_{1\leq i\leq n}|X_{i+1,n} - X_{i,n}|^k \leq S_n(k) \leq \sup_{1\leq i\leq n}|X_{i+1,n} - X_{i,n}|^k.$$

Step 4. We will prove that for each integer $k \geq 1$,

$$\lim_{n\to\infty}\frac{\log[S_n(k)/(S_n(1))^k]}{k\ \log(n-1)} = 0 \qquad \text{a.s.} \tag{3.8}$$

This will be shown to be a consequence of the inequalities (3.9), (3.10) and (3.11) stated below.
We first infer from Step 1, that

$$0 \leq \frac{\log[S_n(k)/(S_n(1))^k]}{k\ \log(n-1)}. \tag{3.9}$$

In view of Step 3, we see that, for each $0 \leq H' < H$, we have with probability one, for sufficiently large values of n,

$$S_n(k) \leq (n-1)^{-H'k},$$

whence, we have for all n large enough

$$\frac{\log(S_n(k)/(n-1)^{-Hk})}{k\ \log(n-1)} \leq H - H'. \tag{3.10}$$

Writing

$$\frac{\log((n-1)^{-Hk}/(S_n(1))^k)}{k\ \log(n-1)} = \frac{-H\ \log(n-1) - \log(S_n(1))}{\log(n-1)} = -H - \frac{\log(S_n(1))}{\log(n-1)}, \tag{3.11}$$

and combining (3.9), (3.10) and (3.11), we obtain that, for all large n,

$$\begin{aligned} 0 \leq \frac{\log(S_n(k)/(S_n(1))^k)}{k\ \log(n-1)} &= \frac{\log(S_n(k)/(n-1)^{-Hk})}{k\ \log(n-1)} + \frac{\log((n-1)^{-Hk}/(S_n(1))^k)}{k\ \log(n-1)} \\ &\leq H - H' - H - \frac{\log(S_n(1))}{\log(n-1)}. \end{aligned}$$

Thus, in view of Step 2, we see that for each $0 \leq H' < H$,

$$0 \leq \liminf_{n\to\infty} \frac{\log(S_n(k)/(S_n(1))^k)}{k\ \log(n-1)} \leq \limsup_{n\to\infty} \frac{\log(S_n(k)/(S_n(1))^k)}{k\ \log(n-1)} \leq H - H'.$$

Since $H' \in [0, H)$, we have

$$\lim_{n\to\infty} \left(\frac{\log(S_n(k)/(S_n(1))^k)}{k\ \log(n-1)} \right) = 0,$$

which is (3.8).

Step 5. By splitting $H_n(k)$ into,

$$H_n(k) = -\frac{\log(S_n(k)/(S_n(1))^k)}{k\ \log(n-1)} - \frac{\log(S_n(1))^k}{k\ \log(n-1)} - \frac{\log\left(\sqrt{\pi}/2^{k/2}\Gamma\left(\frac{k+1}{2}\right)\right)}{k\ \log(n-1)},$$

we therefore complete the proof of (3.7) via Steps 2 and 4. □

Let $\{X_H(t) : 0 \leq t \leq 1\}$ denote a fractional Brownian motion of index H and let $K > 0$ denote a constant. We will introduce two auxiliary sequences of random variables related to X_H. Let $\lfloor x \rfloor$ denote the integer part of x.

For each $n \geq 2$ and $0 \leq i \leq \lfloor \frac{n-1}{K} \rfloor$, set $\mathbb{X}^1_{i,n} = X_H(K\frac{i}{n-1})$. In this case, K may be interpreted as a *time scale factor* .

For each $n \geq 2$ and $0 \leq i \leq (n-1)$, set $\mathbb{X}^2_{i,n} = K^H X_H(\frac{i}{n-1})$. In this case, K^H is the usual *scale factor.*

Set

$$Y^1_{i,n} = (\frac{n-1}{K})^H(\mathbb{X}^1_{i+1,n} - \mathbb{X}^1_{i,n}), \quad 0 \leq i \leq \left\lfloor \frac{n-1}{K} \right\rfloor$$

and

$$Y^2_{i,n} = (\frac{n-1}{K})^H(\mathbb{X}^2_{i+1,n} - \mathbb{X}^2_{i,n}), \quad 1 \leq i \leq n-1.$$

By the strong stationarity of the increments of the fractional Brownian motion, each of these sequences of random variables is stationary in the strong sense. This allows us to denote respectively by $c^1_{0,n}$ and $c^2_{0,n}$ the autocovariance functions of the increments $(Y^1_{i,n}, 1 \leq i \leq \lfloor \frac{n-1}{K} \rfloor)$, $(Y^2_{i,n}, 1 \leq i \leq \lfloor \frac{n-1}{K} \rfloor)$ and, for each $k \geq 1$, by $c^1_{k,n}$, $c^2_{k,n}$ the autocovariance functions of the absolute increments $(|Y^1_{i,n}|^k, 1 \leq i \leq n-1)$, $(|Y^2_{i,n}|^k, 1 \leq i \leq n-1)$.

The following technical lemma shows that the functions $c^1_{0,n}, c^2_{0,n}, c^1_{k,n}$ and $c^2_{k,n}$ do not in fact depend on n, so that we will show that we can denote respectively by c_0 and, for each

$k \geq 1$, by $c_k^\cdot$, the autocovariance functions defined for each $j \geq 1$ and $n > j$, by $c_0(j) = E(Y^1_{1,n}.Y^1_{j+1,n}) = E(Y^2_{1,n}.Y^2_{j+1,n})$ and $c_k(j) = Cov(|Y^1_{1,n}|^k, |Y^1_{j+1,n}|^k) = Cov(|Y^2_{1,n}|^k, |Y^2_{j+1,n}|^k)$.

Lemma 3.3

(1) *For each* $n \geq 2$, *we have* $Y^1_{i,n} \equiv N(0,1)$ *for* $1 \leq i \leq \lfloor \frac{n-1}{K} \rfloor$ *and* $Y^2_{i,n} \equiv N(0,1)$ *for* $1 \leq i \leq n-1$.

(2) *For each* $n, m \geq 2$ *and for all* $1 \leq j < \min(n,m)$ *we have*
(a) $c^1_{0,n}(j) = c^1_{0,m}(j) = c^2_{0,n}(j) = c^2_{0,m}(j)$;
(b) for each $k \geq 1$, $c^1_{k,n}(j) = c^1_{k,m}(j) = c^2_{k,n}(j) = c^2_{k,m}(j)$.

(3) $c_0(j) = O(\frac{1}{j^{2(1-H)}})$ *as* $j \to \infty$.

(4) *For each* $j \geq 1$, *we have* $c_2(j) = 2(c_0(j))^2$.

(5) $c_k(j) = O(c_2(j))$ *as* $j \to \infty$, *for each integer* $k \geq 1$.

Proof.

(1) is trivial.

(2) The proof of this statement will be achieved in two steps, as follows.
Step 1. For each $n \geq 2$ and for $1 \leq j < n$ we have

$$
\begin{aligned}
c^1_{0,n}(j) &= Cov[Y^1_{1,n}, Y^1_{j+1,n}] = \left(\frac{n-1}{K}\right)^{2H} Cov[\mathbb{X}_{2,n} - \mathbb{X}_{1,n}, \mathbb{X}_{j+2,n} - \mathbb{X}_{j+1,n}] \\
&= \left(\frac{n-1}{K}\right)^{2H} Cov[\mathbb{X}_{1,n}, \mathbb{X}_{j+1,n} - \mathbb{X}_{j,n}] \\
&= \left(\frac{n-1}{K}\right)^{2H} \{Cov[\mathbb{X}_{1,n}, \mathbb{X}_{j+1,n} - \mathbb{X}_{1,n}] - Cov[\mathbb{X}_{1,n}, \mathbb{X}_{j,n} - \mathbb{X}_{1,n}]\} \\
&= \frac{(\frac{n-1}{K})^{2H}}{2}\{Var(\mathbb{X}_{j+1,n}) - 2Var(\mathbb{X}_{j,n}) + Var(\mathbb{X}_{j-1,n})\} \\
&= \frac{1}{2}\{(j+1)^{2H} - 2j^{2H} + (j-1)^{2H}\}.
\end{aligned}
$$

The same result holds for $c^2_{0,n}(j)$, whence (2)(a) in the statement of the lemma.
Step 2. For each integer $k \geq 2$, for each $n \geq 2$, for all $1 \leq j < n$ we have

$$
c^1_{k,n}(j) = Cov[|Y^1_{1,n}|^k, |Y^1_{j+1,n}|^k] = E(|Y^1_{1,n}|^k |Y^1_{j+1,n}|^k) - E(|Y^1_{1,n}|^k)E(|Y^1_{j+1,n}|^k).
$$

By (1) and Step 1 $(Y^1_{1,n}, Y^1_{j+1,n})$ and $(Y^2_{1,n}, Y^2_{j+1,n})$ have the same bivariate normal distribution which is independent of n. Hence, we obtain statement (2)(b), as sought.

(3) The statement (2) (a) yields that the function c_0 is well defined. By Step 1 of the proof of statement (2), for each $j > 1$ we have

$$
\begin{aligned}
c_0(j) &= \frac{1}{2}j^{2H}[1 + \frac{2H}{j} + \frac{H(2H-1)}{j^2} - 2 + 1 - \frac{2H}{j} + \frac{H(2H-1)}{j^2} + o(\frac{1}{j^2})] \\
&= \frac{H(2H-1)}{j^{2(1-H)}} + o(\frac{1}{j^{2(1-H)}}) \\
&= O(\frac{1}{j^{2(1-H)}}), \quad \text{as } j \to \infty,
\end{aligned}
$$

which is the result we seek.

(4) The statement (2) (b) yields that for each $k \geq 1$, the function c_k is well defined. For the sake of notational simplicity for each $j \geq 1$ we denote by $Y_j = Y^1_{j,n}$ so that $c_k(j) = Cov(|Y_1|^k.|Y_{j+1}|^k)$.
We recall that *if* (X, Y) *follows a bivariate normal distribution with means* $E(X) = E(Y) = 0$*, variances* $Var(X) = Var(Y) = 1$ *and correlation* $Corr(X, Y) = \rho$*, then the conditonal distribution of* X *given* $Y = y$ *is the density function of a normal* $N(\rho y, 1 - \rho^2)$ *law* (see e.g. Grimmet *et al* (1992) p.384).
For each $j \geq 1$, $c_2(j) = Cov[Y_1^2, Y_{j+1}^2] = E(Y_1^2 Y_{j+1}^2) - E(Y_1^2)E(Y_{j+1}^2) = E(Y_1^2 Y_{j+1}^2) - 1$.
By Step 2 of the proof of statement (2), and since the correlation between $Y^1_{1,n}$ and $Y^1_{j+1,n}$ equals $c_0(j)$, we have $E(Y_{j+1}^2 \mid Y_1) = Var(Y_{j+1}^2 \mid Y_1) + (E(Y_{j+1} \mid Y_1))^2 = 1 - (c_0(j))^2 + (c_0(j))^2 Y_1^2$.
Thus we have $c_2(j) = E[Y_1^2.E(Y_{j+1}^2 \mid Y_1)] - 1 = (1 - (c_0(j))^2)E(Y_1^2) + (c_0(j))^2 E(Y_1^4) - 1 = 2(c_0(j))^2$, which proves statement (4).

(5) We keep the notation of statement (4). For each even integer $k \geq 2$, and integer $j \geq 1$, we infer from Step 2 of the proof of statement (2) that

$$\begin{aligned} c_k(j) &= E[(Y_1^k.E(Y_{j+1}^k \mid Y_1)] - (E(Y_1^k))^2 \\ &= E[Y_1^k.E((Y_{j+1} - c_0(j)Y_1 + c_0(j)Y_1)^k \mid Y_1)] - (E(Y_1^k))^2 \\ &= E[Y_1^k.E(\sum_{i=0}^{k} \tbinom{k}{i}(Y_{j+1} - c_0(j)Y_1)^i (c_0(j)Y_1)^{k-i} \mid Y_1)] - (E(Y_1^k))^2 \\ &= E[Y_1^k \sum_{i=0}^{k} \tbinom{k}{i}(c_0(j)Y_1)^{k-i} E((Y_{j+1} - c_0(j)Y_1)^i \mid Y_1)] - (E(Y_1^k))^2. \end{aligned}$$

If i is an odd integer we have $E((Y_{j+1} - c_0(j)Y_1)^i \mid Y_1) = 0$, therefore we may limit ourselves to the case of an even i, in which case we have for all $j \geq 1$

$$c_k(j) = E[Y_1^k \sum_{i=0}^{k/2} \tbinom{k}{2i}(c_0(j)Y_1)^{k-2i} E((Y_{j+1} - c_0(j)Y_1)^{2i} \mid Y_1)] - (E(Y_1^k))^2.$$

The conditional law of Y_{j+1} given Y_1 is a normal distribution with mean $c_0(j)Y_1$ and variance $1 - c_0^2(j)$. Thus, using (3.1) we have

$$c_k(j) = \sum_{i=0}^{k/2} \tbinom{k}{2i} E[Y_1^k (c_0(j)Y_1)^{k-2i} E(|Y_1|^{2i})(1 - c_0^2(j))] - (E(Y_1^k))^2.$$

It follows that for $i \in [0, k/2)$, as $j \to \infty$

$$E[Y_1^k (c_0(j)Y_1)^{k-2i} E(|Y_1|^{2i})(1 - c_0^2(j))] = O(c_0^2(j)),$$

whereas for $i = k/2$,

$$E[Y_1^k E(|Y_1|^k)(1 - c_0^2(j))] = [E(Y_1^k)]^2 - c_0^2(j).[E(Y_1^k)]^2.$$

Combining these two cases, we see that $c_k(j) = O(c_2(j))$ as $j \to \infty$.
Assume next that the integer $k \geq 1$ is odd. We have the following two steps.
Step 1. Let $X \equiv N(m, \sigma^2)$, $Y = X - m$, and denote by f_X and f_Y the density

functions of X and Y respectively. We have the chain of equalities

$$E(|X|^k) = \int_{-\infty}^{0} -x^k f_X(x)dx + \int_0^m x^k f_X(x)dx + \int_m^\infty x^k f_X(x)dx$$

$$= \sum_{l=0}^{k} \binom{k}{l} \int_{-\infty}^{0} (m-x)^l(-m)^{k-l} f_X(x)dx + \int_0^m x^k f_X(x)dx + \int_0^\infty (y+m)^k f_Y(y)dy$$

by the change of variables $y = x - m$ in the last integral. Thus

$$E(|X|^k) = \sum_{l=0}^{k} \binom{k}{l} \int_m^\infty y^l(-m)^{k-l} f_Y(y)dy + \int_0^m x^k f_X(x)dx + \sum_{l=0}^{k} \binom{k}{l} \int_0^\infty y^l m^{k-l} f_Y(y)dy$$

by the change of variables $y = m - x$ in the first integral. Therefore

$$E(|X|^k) = 2\sum_{l=1}^{(k+1)/2} \binom{k}{l} \int_m^\infty y^{2l-1} m^{k-2l+1} f_Y(y)dy - \sum_{l=0}^{k} \binom{k}{l} \int_0^m y^l m^{k-l} f_Y(y)dy + \int_0^m x^k f_X(x)dx.$$

Since for each $0 \le l' \le k$, $\quad 0 \le y \le m$ we have $|y^{l'} m^{k-l'} f_Y(y)| \le |m|^k f_Y(0)$, then we easily obtain

$$\begin{aligned} E(|X|^k) &= \sum_{l=1}^{(k+1)/2} \binom{k}{l} m^{k-2l+1} E(|Y|^{2l-1}) + O(m^2) \quad \text{as } m \to 0 \\ &= E(|Y|^k) + O(m^2) \quad \text{as } m \to 0. \end{aligned}$$

Step 2. For each $j \ge 1$, we have

$$\begin{aligned} c_k(j) &= Cov[|Y_1|^k, |Y_{j+1}|^k] = E(|Y_1|^k |Y_{j+1}|^k) - E(|Y_1|^k)E(|Y_{j+1}|^k) \\ &= E(|Y_1|^k |Y_{j+1}|^k) - (E(|Y_1|^k))^2 = E(|Y_1|^k . E(|Y_{j+1}|^k \mid Y_1)) - (E(|Y_1|^k))^2. \end{aligned}$$

The conditional law of Y_{j+1} given Y_1 is a normal distribution with mean $c_0(j)Y_1$ and variance $1 - c_0^2(j)$. Thus, using (3.1) it follows from Step 1 above that

$$\begin{aligned} c_k(j) &= E[|Y_1|^k . (E(|Y_1|^k) + O((E(|Y_1|^k))^2 (c_0(j))^2))] - (E(|Y_1|^k))^2 \\ &= O((c_0(j))^2) \quad \text{as } j \to \infty. \end{aligned}$$

This concludes our proof. □

We denote by $u_n \approx v_n$ (resp. $u_n \sim v_n$) the fact that $u_n/v_n \to \lambda$ as $n \to \infty$ for some $\lambda \in (0, \infty)$ (resp. for $\lambda = 1$) .

Lemma 3.4 *For each $x \in \mathbb{R}$, we have*

$$\frac{1}{n^x}\sum_{i=1}^{n}\frac{1}{i} \approx \frac{\log n}{n^x},$$

$$-\infty < y < 1, \quad \frac{1}{n^x}\sum_{i=1}^{n}\frac{1}{i^y} \approx \frac{1}{n^{x+y-1}},$$

$$y > 1, \quad \frac{1}{n^x}\sum_{i=1}^{n}\frac{1}{i^y} \approx \frac{1}{n^x}.$$

Proof. The results are easily obtained through elementary analysis. □

For each $k \ge 1$ we set $m_k = E(|Y|^k)$ and $v_k = Var(|Y|^k)$ where Y is a normal $N(0, 1)$ random variable. Fix any $0 < b_H < \min\{\frac{H}{1-H}, \frac{3}{4}\}$ and denote by $(\nu_n)_{n \ge 1}$ be the sequence defined by $\nu_n = \lfloor n^{\frac{1}{1-H} - b_H} \rfloor$.

Lemma 3.5 *We have for each integer $k \geq 1$,*

$$\frac{1}{n-1} \left| \sum_{i=1}^{n-1} (|Y_{i,n}^1|^k - m_k) \right| \xrightarrow[n\to\infty]{P} 0, \tag{3.12}$$

$$\frac{1}{n-1} \left| \sum_{i=1}^{n-1} (|Y_{i,n}^2|^k - m_k) \right| \xrightarrow[n\to\infty]{P} 0,$$

and

$$\lim_{n\to\infty} \frac{1}{\nu_n - 1} \sum_{i=1}^{\nu_n - 1} |Y_{i,\nu_n}^1|^k = \lim_{n\to\infty} \frac{1}{\nu_n - 1} \sum_{i=1}^{\nu_n - 1} |Y_{i,\nu_n}^2|^k = \frac{2^{k/2}\Gamma(\frac{k+1}{2})}{\sqrt{\pi}} \quad a.s. \tag{3.13}$$

Proof. Set for convenience $Y_{i,n} = Y_{i,n}^1$ for each $1 \leq i \leq n-1$. Without loss of generality we may limit ourselves to establish the lemma for the sequence $(Y_{i,n})_{1\leq i\leq n-1}$.
Let $(\mu_n)_{n\geq 1}$ be an increasing sequence of integers.
Fix any $\epsilon > 0$ and set $p_n = P\left(\frac{1}{\mu_n - 1} \left| \sum_{i=1}^{\mu_n - 1} (|Y_{i,\mu_n}|^k - m_k) \right| > \epsilon\right)$.
We have

$$\begin{aligned}
p_n &\leq \frac{1}{(\mu_n - 1)^2 \epsilon^2} E\left(\left| \sum_{i=1}^{\mu_n - 1} (|Y_{i,\mu_n}|^k - m_k) \right|^2 \right) \\
&= \frac{1}{(\mu_n - 1)^2\epsilon^2} \{ \sum_{i=1}^{\mu_n - 1} Var(|Y_{i,\mu_n}|^k) + 2 \sum_{1\leq i<j\leq \mu_{n-1}} Cov(|Y_{i,\mu_n}|^k, |Y_{j,\mu_n}|^k) \} \\
&= \frac{1}{(\mu_n - 1)^2\epsilon^2} \{ (\mu_n - 1)v_k + 2 \sum_{i=1}^{\mu_n - 2} (\mu_n - i - 1) c_k(i) \} \\
&= \frac{v_k}{(\mu_n - 1)\epsilon^2} + \frac{2}{(\mu_n - 1)\epsilon^2} \sum_{i=1}^{\mu_n - 2} c_k(i) - \frac{2}{(\mu_n - 1)^2\epsilon^2} \sum_{i=1}^{\mu_n - 2} i c_k(i).
\end{aligned}$$

By Lemma 3.3 we obtain that
$p_n \leq \frac{v_k}{(\mu_n - 1)\epsilon^2} + \frac{1}{\mu_n - 1} \sum_{i=1}^{\mu_n - 2} O(\frac{1}{i^{4(1-H)}}) - \frac{1}{(\mu_n - 1)^2} \sum_{i=1}^{\mu_n - 2} O(\frac{1}{i^{3-4H}})$.
In view of Lemma 3.4 we are led to the following three cases.

- $0 < H < \frac{1}{2}$
  $p_n \leq \frac{v_k}{\mu_n \epsilon^2} + O(\frac{1}{\mu_n}) + O(\frac{1}{\mu_n^2})$.
  In which case, $p_n = O(\frac{1}{\mu_n})$.
- $\frac{1}{2} \leq H \leq \frac{3}{4}$
  $p_n \leq \frac{v_k}{\mu_n \epsilon^2} + O(\frac{\log n}{\mu_n}) + O(\frac{1}{\mu_n^2 \mu_n^{2-4H}})$.
  In which case, $p_n = O(\frac{\log \mu_n}{\mu_n})$;
- $\frac{3}{4} < H < 1$
  $p_n \leq \frac{v_k}{\mu_n \epsilon^2} + O(\frac{1}{\mu_n \mu_n^{3-4H}}) + O(\frac{1}{\mu_n^2 \mu_n^{2-4H}})$.
  In which case, $p_n = O(\frac{1}{\mu_n^{4(1-H)}})$.

When $\mu_n = n$ we obtain (3.12), whereas when $\mu_n = \nu_n$ we obtain (3.13) by an application of Borel-Cantelli lemma. □

Remark 3.6 Formula (3.12) in the previous lemma gives a result which holds in probability. We conjecture that the corresponding statements hold with almost sure convergence.

This is obviously true when $H = 1/2$. However the proof of the latter statement when $H \neq 1/2$ needs to be much more involved than the present one.

The following theorem establishes a remarkable property of fractional Brownian motion. First, we recall some notation. Let $F = \{(t, X_H(t)) : t \in [0,1]\}$ denote the graph of a fractional Brownian motion, with fractal dimension $s = 2 - H$. Let further $K > 0$ be defined as in Proposition 3.1 and Lemma 3.3 such that K^H is the usual scale factor. For $i = 1 \dots 5$, we denote by $\underline{c}^{(i)}$ and $\overline{c}^{(i)}$ be the random variables defined almost surely by

$$\underline{c}^{(i)} = \varliminf_{\delta \to 0} \delta^s N_\delta(F) \quad \text{and} \quad \overline{c}^{(i)} = \varlimsup_{\delta \to 0} \delta^s N_\delta(F),$$

where $N_\delta^{(i)}(F)$ is respectively one of the following alternative definitions:

(1) For $i = 1$, the number of δ-mesh squares that intersect F;
(2) For $i = 2$, the smallest number of squares of size δ that cover F;
(3) For $i = 3$, the smallest number of closed balls of diameter δ that cover F;
(4) For $i = 4$, the smallest number of sets of diameter δ that cover F;
(5) For $i = 5$, the largest number of disjoint balls of diameter δ with centres in F.

Theorem 3.7 *We have with probability one*

(1) $\underline{c}^{(1)} = \overline{c}^{(1)} = K^H \sqrt{\frac{2}{\pi}}$;
(2) $\frac{K^H}{\sqrt{2\pi}} \leq \underline{c}^{(2)} \leq \overline{c}^{(2)} \leq K^H \sqrt{\frac{2}{\pi}}$;
(3) *For* $i = 3, 4$ *and* 5 $\frac{K^H}{\sqrt{2\pi}} \leq \underline{c}^{(i)} \leq \overline{c}^{(i)} \leq K^H \frac{2}{\sqrt{\pi}}$.

Proof.

(1) As in Lemma 3.5 we write $\nu_n = \lfloor n^{\frac{1}{1-H} - b_H} \rfloor$, and choose $0 < b_H < \min\{\frac{H}{2(1-H)}, \frac{3}{4}\}$. We now make use of the notation and proof of Lemma 2.2 with $X = X_H$. We have seen that $L_{\nu_n} \leq N_{\delta_{\nu_n}, \nu_n} \delta_{\nu_n} \leq (L_{\nu_n} + 1)$ and that $N_{\delta_{\nu_n}, \nu_n} = (1 + o(1)) N_{\delta_{\nu_n}}$. This yields

$$L_{\nu_n} \delta_{\nu_n}^{1-H} \leq N_{\delta_{\nu_n}, \nu_n} \delta_{\nu_n}^s \leq (L_{\nu_n} + 1) \delta_{\nu_n}^{1-H},$$

or equivalently with the notation of Lemma 3.3

$$\frac{1}{\nu_n - 1} \sum_{i=1}^{\nu_n - 1} |Y_{i,\nu_n}^2| K^H \leq N_{\delta_{\nu_n}, \nu_n} \delta_{\nu_n}^s \leq \frac{1}{\nu_n - 1} \sum_{i=1}^{\nu_n - 1} |Y_{i,\nu_n}^2| K^H + \delta_{\nu_n}^{1-H}.$$

Thus, in this particular case, (1) holds by application of Lemma 3.5 with $k = 1$.

Next, we will show that (1) holds for any $\delta > 0$. For any $\varepsilon > 0$, there exists $N_1 > 0$ such that, for each $n > N_1$,

$$|N_{\delta_n} \delta_n^s - c^{(1)}| \leq \varepsilon/6, \tag{a}$$

and there exists an $N_2 > 0$ such that, for each $n > N_2$,

$$|N_{\delta_n} \delta_n^s - N_{\delta_{n+1}} \delta_{n+1}^s| \leq \varepsilon/6. \tag{b}$$

We conclude by showing that there exists an $N_3 > 0$ such that, for each $n > N_3$,

$$N_{\delta_{n+1}} (\delta_n - \delta_{n+1}) \leq \varepsilon/6. \tag{c}$$

To prove this last inequality, the following arguments are needed.

The study of the variations of the function $\phi(\eta) = (H-\eta)(H-\eta-b_H(1-\eta)(1-H))$ around $H=\eta$ shows that there exists η such that $0<\eta<H<\eta+b_H(1-\eta)(1-H)$. Hence, by Lemma 2.1 and definition (2.0), there exists $N_3>0$ such that, for each $n>N_3$, we have

$$\frac{\log(N_{\delta_{n+1}})}{-\log(\delta_{n+1})} \leq 2-\eta.$$

Therefore, as $n\to\infty$, we have

$$\begin{aligned}
N_{\delta_{n+1}}(\delta_n-\delta_{n+1}) &\leq \frac{\delta_n-\delta_{n+1}}{\delta_{n+1}^{2-\eta}} \\
&\sim \frac{n^{b_H-\frac{1}{1-H}-1}}{n^{(b_H-\frac{1}{1-H})(2-\eta)}} \times (\frac{1}{1-H}-b_H) \\
&= \frac{1}{n^{1+(b_H-\frac{1}{1-H})(1-\eta)}} \times (\frac{1}{1-H}-b_H) \\
&= \frac{1}{n^{\frac{1-H+b_H(1-H)(1-\eta)-1+\eta}{1-H}}} \times (\frac{1}{1-H}-b_H) \\
&= \frac{1}{n^{\frac{\eta-H+b_H(1-H)(1-\eta)}{1-H}}} \times (\frac{1}{1-H}-b_H) \longrightarrow 0.
\end{aligned}$$

which suffices for our needs.

By combining the above statements (a), (b) and (c), we see that, for sufficiently small values of $\delta>0$, there exists $n>max\{N_1,N_2,N_3\}$, for which $\delta_{n+1}<\delta<\delta_n$, and,

$$\begin{aligned}
|N_\delta\delta^s-c^{(1)}| &\leq |N_\delta\delta_n^s-c^{(1)}|+|N_\delta\delta_{n+1}^s-c^{(1)}| \\
&\leq |N_{\delta_n}\delta_n^s-c^{(1)}|+|N_{\delta_{n+1}}\delta_n^s-c^{(1)}|+|N_{\delta_n}\delta_{n+1}^s-c^{(1)}|+|N_{\delta_{n+1}}\delta_{n+1}^s-c^{(1)}| \\
&\leq 2|N_{\delta_n}\delta_n^s-c^{(1)}|+2|N_{\delta_{n+1}}\delta_{n+1}^s-c^{(1)}|+N_{\delta_n}(\delta_n^s-\delta_{n+1}^s)+N_{\delta_{n+1}}(\delta_n^s-\delta_{n+1}^s) \\
&\leq \varepsilon.
\end{aligned}$$

(2) Recall the definition of $N_\delta^{(1)}(F)$ and $N_\delta^{(2)}(F)$. With the notation of Lemma 2.2, by choosing $X=X_H$ in this lemma, and defining X_n accordingly, we denote respectively by $N_{\delta_n}^{1,n}(F)$ and $N_{\delta_n}^{2,n}(F)$ the number of δ_n-mesh squares and the smallest number of squares of size δ that intersect the graph of X_n. We will establish the following inequalities

$$N_{\delta_n}^{1,n}(F) \geq N_{\delta_n}^{2,n}(F) \geq \frac{1}{2}N_{\delta_n}^{1,n}(F). \tag{3.14}$$

The first inequality in (3.14) is obvious. For the second inequality we assume without loss of generality that $(n-1)/2\in\mathbb{N}$. For $1\leq i\leq(n-1)/2$, we denote by F_i, the graph of X_n defined on $[\frac{2i-1}{n},\frac{2i+1}{n}]$. We have the obvious equality

$$N_{\delta_n}^{1,n}(F) = \sum_{i=1}^{(n-1)/2} N_{\delta_n}^1(F_i).$$

We show in Annex A that the number N_i of any set of squares of size δ covering F_i satisfies the inequality

$$N_i \geq \frac{1}{2}N_{\delta_n}^1(F_i). \tag{3.15}$$

By (3.15), we have

$$\sum_{i=1}^{(n-1)/2} N_i \geq \frac{1}{2} N^1_{\delta_n}(F_i),$$

which completes the proof of the inequalities in (3.14).

Since X_n converges uniformly to X on $[0,1]$ as $n \to \infty$, we have $N^{1,n}_{\delta_n}(F) \sim N^1_\delta(F)$ and $N^{2,n}_\delta(F) \sim N^2_\delta(F)$. This, combined with (1) and (3.14) prove (2).

(3) The proof of the first inequality is obtained by the same arguments as the ones used for (2). The last inequality is involved by the fact that *any square of size δ is covered by a closed ball of diameter $\sqrt{2}\delta$.* □

Proposition 3.8 *Under the assumptions of Proposition 3.1 and Lemmas 3.3 and 3.5, for each integer $k \geq 1$, we have*

$$\frac{S_n(k)}{E_n(k)} \xrightarrow[n\to\infty]{P} 1, \qquad (\log n)(H_{n,K}(k) - H) \xrightarrow[n\to\infty]{P} 0, \tag{3.15}$$

and with the notation of Proposition 3.2,

$$\lim_{n\to\infty} \left(\frac{S_{\nu_n}(k)}{E_{\nu_n}(k)}\right) = 1 \quad a.s., \qquad \lim_{n\to\infty} (\log \nu_n)(H_{\nu_n,K}(k) - H) = 0 \quad a.s. \tag{3.16}$$

Proof. By Propositions 3.1 and 3.2 we have

$$\frac{S_n(k)}{E_n(k)} = \left(\frac{n-1}{K}\right)^{H - H_n(k)}$$

Lemma 3.5 in combination with Slutsky's theorem (see e.g. Grimmett *et al* (1992) p.285) completes the proof. □

Let $m \geq 1$ be an integer. The following proposition establishes the consistency of the new class of estimators defined by

$$\hat{H}_n(k) = \frac{1}{k\,(\log m)} \log\left(\frac{S^m_{p_n}(k)}{S^1_n(k)}\right)$$

where $p_n = \lfloor \frac{n}{m} \rfloor$,

$$S^1_n(k) = S_n(k) = \frac{1}{n-1}\sum_{i=1}^{n-1} |X_{i+1,n} - X_{i,n}|^k \quad \text{and} \quad S^m_{p_n}(k) = \frac{1}{p_n - 1}\sum_{i=1}^{p_n-1} |X^n_{m(i+1)} - X^n_{mi}|^k.$$

With the notation of Lemma 3.5 the following proposition holds.

Proposition 3.9 *For each integer $k \geq 1$, $\hat{H}_n(k) \xrightarrow[n\to\infty]{P} H$ and $\lim_{n\to\infty} \hat{H}_{\nu_n}(k) = H$ a.s.*

Proof. We recall Proposition 3.8 and the notation of Lemma 3.3 where the sequence $(\mathbb{X}^1_{i,n})_{1\leq i\leq n}$ is defined. Setting $K = 1$, there exists $c(k) > 0$ such that

$$S^1_{\nu_n}(k) = (1 + o(1))c(k)\nu_n{}^{-Hk} \qquad \text{as } n \to \infty.$$

With the notation of Lemma 3.3 we consider next the sequence $(\mathbb{X}^1_{i,n})_{1\le i\le n}$, with $K = m$. We have

$$S^m_{p_{\nu_n}}(k) = (1+o(1))c(k){p_{\nu_n}}^{-Hk} \qquad \text{as } n\to\infty.$$

We thus conclude that $\lim_{n\to\infty} \hat{H}_{\nu_n}(k) = H$ a.s.
We note further that, for any $\epsilon > 0$,

$$P[\|(\frac{S^1_n(k)}{E_n(k)}, \frac{S^1_{p_n}(k)}{E_{p_n}(k)}) - (1,1)\|_2 > \epsilon] = P[\left(\frac{S^1_n(k)}{E_n(k)} - 1\right)^2 + \left(\frac{S^1_{p_n}(k)}{E_{p_n}(k)} - 1\right)^2 > \epsilon^2]$$

$$\le P[\left(\frac{S^1_n(k)}{E_n(k)} - 1\right)^2 > \frac{\epsilon^2}{2}] + P[\left(\frac{S^1_{p_n}(k)}{E_{p_n}(k)} - 1\right)^2 > \frac{\epsilon^2}{2}] \longrightarrow 0,$$

where we have made use of the arguments in the proof of Lemma 3.5. Thus,

$$(\frac{S^1_n(k)}{E_n(k)}, \frac{S^1_{p_n}(k)}{E_{p_n}(k)}) \xrightarrow[n\to\infty]{P} (1,1),$$

so that an argument based on Slutsky's theorem complete the proof of Proposition 3.9. □

Remark 3.10 Choices of $K = m$ which are convenient for practical applications are $K = 2$, 3 or 4. The corresponding estimators appear to perform as well for generalized fractional Bronian motion (*i.e.* for $K^H > 0$) as for standard fractional Brownian motion (*i.e.* for $K^H = 1$).

4. Rate of convergence. - Asymptotic Confidence intervals. - Test of fit to fractional Brownian motion

In this section, we will investigate the asymptotic behavior of our estimators. The rates of convergence depend upon the parameters H and k. The main results of this section are stated in Proposition 4.2 (where we obtain the rate of convergence of our estimators), Proposition 4.4 (where we propose a statistic which provides a possible test of fit to a fractional Brownian motion), and Proposition 4.6 (where we describe the asymptotic behavior of this statistic). We start by a simple lemma.

Lemma 4.1 *Let $(Z_n)_{n\in\mathbb{N}}$, $(Y_n)_{n\in\mathbb{N}}$ be sequences of random variables and Z a random variable such that: $Z_n \to Z$ in distribution and $Z_n = (1+o_P(1))Y_n$ as $n\to 0$. Then $Y_n \to Z$ in distribution.*

Proof. We recall (see e.g. Grimmet *et al* (1992) p.277;p.285) that if V_n and W_n, $n\ge 1$ are sequences of random variables then

- if V_n tends in distribution to $a\in\mathbb{R}$ then V_n tends to a in probability.
- if V_n tends in distribution to $a\in\mathbb{R}$ and W_n tends to the random variable W in distribution then $V_n + W_n$ tends in distribution to $a + W$.
- if V_n tends in distribution to $a\in\mathbb{R}$ and W_n tends to the random variable W in distribution then V_nW_n tends in distribution to aW.

Our assumption entails that $\left|\frac{Y_n-Z_n}{Z_n}\right| \xrightarrow[n\to\infty]{P} 0$ so that $\left|\frac{Y_n-Z_n}{Z_n}\right| \to 0$ in distribution. It follows therefore that $\left|\frac{Z(Y_n-Z_n)}{Z_n}\right| \to 0$ in distribution. By Slutsky's theorem we have $\left|\frac{Z_n}{Z}\right| \xrightarrow[n\to\infty]{P} 1$

so that $\left|\frac{Z_n}{Z}\right| \to 1$ in distribution. Thus, $|Y_n - Z_n| = \left|\frac{Z(Y_n - Z_n)}{Z_n}\right| . \left|\frac{Z_n}{Z}\right| \to 0$ in distribution. Finally, we have $Y_n = (Y_n - Z_n) + Z_n$, $Y_n - Z_n \to 0$ in distribution and $Z_n \to Z$ in distribution, whence $Y_n \to Z$ in distribution. □

The following two theorems will be used in the sequel. In the first place, we recall a result of Dobrushin and Major (1979) which we state in theorem A below, in a simplified version.

Let $\{Y_n : n = 0, \pm 1, \dots\}$ be a stationnary Gaussian sequence with $EY_n = 0$, $E(Y_n^2) = 1$. We assume that the correlation function $c_0(n) = E(Y_0 Y_n)$ satisfies the relation $c_0(n) = n^{-\alpha} L(n)$, $0 < \alpha < 1$, where $L(t)$, $t \in (0, \infty)$, is a slowly varying function [i.e. $\lim_{s\to\infty} \frac{L(st)}{L(s)} = 1$ for every $t \in (0,\infty)$, and $L(t)$ is integrable on every finite interval]. We consider a real function $\mathbb{H}(y)$ such that $\mathbb{H}(y)$ does not vanish on a set of positive measure, $\int_{-\infty}^{\infty} \mathbb{H}(y) exp\left(-\frac{y^2}{2}\right) dy = 0$, and $\int_{-\infty}^{\infty} [\mathbb{H}(y)]^2 exp\left(-\frac{y^2}{2}\right) dy < \infty$. We shall consider the following expansion of $\mathbb{H}(y)$: $\mathbb{H}(y) = \sum_{j=1}^{\infty} a_j H_j(y)$ with $\sum_{j=1}^{\infty} a_j^2 j! < \infty$, where H_j is the j-th Hermite polynomial with leading coefficient 1.

Theorem A. *With the notation and assumptions above, assume that* $\alpha < \frac{1}{l}$, *where* $l = \min\{j \geq 1 : a_j \neq 0\}$.
Set $A_n = n^{1-\frac{l\alpha}{2}} |L(n)|^{\frac{l}{2}}$, *then* $T_n = \frac{1}{A_n} \sum_{j=1}^{n-1} \mathbb{H}(Y_j) \xrightarrow{d} T$ *as* $n \to \infty$, *where*

$$T = \left(2\Gamma(\alpha) cos(\frac{\alpha\pi}{2})\right)^{-\frac{l}{2}} a_l \int \cdots \int e^{i(y_1 + \dots + y_l)} \frac{e^{i(y_1+\dots+y_l)} - 1}{i(y_1 + \dots + y_l)} |y_1|^{\frac{\alpha-1}{2}} \dots |y_l|^{\frac{\alpha-1}{2}} dW(y_1) \dots dW(y_l). \tag{4.1}$$

In (4.1), the integral is meant as a multiple Wiener-Itô integral with respect to the random spectral measure W *of the white-noise process.*

In the second place, we will make use of a theorem of Breuer and Major (1983) which we state below in a simplified form.

Theorem B. *With the notation and assumptions above, assume that* $\sum_{n \in \mathbb{N}} |c_0(n)|^l < \infty$, *where* $l = \min\{j \geq 1 : a_j \neq 0\}$. *Set* $A_n^* = n^{1/2}$ *and* $T_n^* = \frac{1}{A_n^*} \sum_{j=1}^{n-1} \mathbb{H}(Y_j)$. *Then* $\lim_{n\to\infty} E((T_n^*)^2) = \sigma^2 < \infty$ *and* $T_n^* \xrightarrow{d} \sigma T^*$ *as* $n \to \infty$, *where* T^* *denotes a standard normal* $N(0,1)$ *random variable.*

We establish the rates of convergence of our estimators in the next proposition.

Proposition 4.2 *With the notation of Propositions 3.1 and 3.2, for a fractional Brownian motion* X_H, *we have for each integer* $k \geq 1$:
Whenever $H < \frac{3}{4}$ *there exists a* $V_H(k) > 0$, *such that*

$$\sqrt{n} \log(n/K)(H_{n,K}(k) - H) \xrightarrow{d} N(0, V_H(k)) \tag{4.2}$$

and whenever $H > \frac{3}{4}$ *we have*

$$n^{2(1-H)}\log(n/K)(H_{n,K}(k) - H) \xrightarrow{d} \frac{k.\lambda}{2} \iint_{\mathbb{R}^2} e^{i(y_1+y_2)} \frac{e^{i(y_1+y_2)} - 1}{i(y_1+y_2)} |y_1|^{\frac{\alpha-1}{2}} |y_2|^{\frac{\alpha-1}{2}} dW(y_1)dW(y_2) \tag{4.3}$$

where $\alpha = 2(1-H)$, $\lambda = \frac{H(2H-1)}{2\Gamma(\alpha)cos(\frac{\alpha\pi}{2})}$ *and* W *is the spectral mesure of the white-noise process.*

Proof. We need the three following steps.
Step 1. In this step, we prove that $\frac{S_n(k)}{E_n(k)} - 1 = (1 + o_P(1))\log(n/K)(H - H_{n,K}(k))$. Towards this we aim, we use (3.1) and Proposition 3.2 to obtain that $\frac{S_n(k)}{E_n(k)} - 1 = (\frac{n-1}{K})^{H-H_{n,K}(k)} - 1 = e^{(H-H_{n,K}(k))\log(\frac{n-1}{K})} - 1$. The result follows by Proposition 3.8 and Slutsky's theorem.

Step 2. For each integer $k \geq 1$, set

$$\mathbb{H}^k(y) = |y|^k - E_k, \tag{4.4}$$

with $E_k = E(|Y|^k)$ where Y is a standard normal $N(0,1)$ random variable. For each $k \geq 1$, $\mathbb{H}^k$ obviously satisfies the conditions needed for the application of Theorems A and B. Write $\mathbb{H}^k(y) = \sum_{j=0}^{\infty} a_j^k H_j(y)$ where the Hermite polynomial $H_j(x)$ of degree j is defined by the relation (see e.g. Taqqu (1975))

$$\left(\frac{d}{dx}\right)^j e^{-x^2/2} = (-1)^j H_j(x) e^{x^2/2}.$$

We will establish statements (4.5) and (4.6) below.

$$|y|^k - E_k = \frac{k}{2}E_k H_2(y) + \frac{k(k-2)}{24}E_k H_4(y) + \sum_{j=3}^{\infty} a_{2j}^2 H_{2j}(y), \tag{4.5}$$

and

$$a_2^2 = 1, \text{ and for each } j \neq 2,\ a_j^2 = 0. \tag{4.6}$$

It is easy to show that $H_0(x) = 1$, $H_1(x) = x$, $H_2(x) = x^2 - 1$, $H_3(x) = x^3 - 3x$ and $H_4(x) = x^4 - 6x^2 + 3$. It is well known that the sequence $(H_j(x)/\sqrt{j!})_{j\geq 0}$ forms a complete orthonormal system of the Hilbert space with respect to the normal density Φ, $L^2(\mathbb{R}, \Phi)$. Since the functions $(H_{2j})_{j\geq 0}$ are even and $(H_{2j+1})_{j\geq 0}$ are odd, it easily follows that, for each $j \geq 0$, $a_{2j+1}^k = 0$. The following classical fomula for each $j \geq 0$,

$$a_j^k = \frac{1}{j!}\int_{\mathbb{R}} \mathbb{H}^k(x) H_j(x) \Phi(x) dx,$$

then gives without difficulty (4.5). We have immediately $\mathbb{H}^2 = H_2$, which yields (4.6).

Step 3. We now apply Theorems A and B together with (4.4) and (4.5). On the one hand, we have

$$l = 2 = \min\{j \geq 1, a_j^k \neq 0\} \quad \text{and} \quad a_l^k = \frac{k}{2}E_k. \tag{4.7}$$

and from the proof of Lemma 3.3 (3) we have $c_0(j) = \frac{H(2H-1)}{j^{2(1-H)}} + o(\frac{1}{j^{2(1-H)}}) = j^{-\alpha}L(j)$ as $j \to \infty$, with $\alpha = 2(1-H)$ and $L(j) \sim H(2H-1)$. Hence $L(t)$, $t \in (0,\infty)$, is obviously a slowly varying function. We achieve our proof by considering two cases.

Case A. The condition $\alpha = 2(1-H) < 1/l$ is equivalent to $H > 3/4$. Set $A_n = n^{1-\alpha}L(n)$ and $T_n = \frac{1}{A_n}\sum_{j=1}^{n-1} \mathbb{H}^k(Y^2_{j,n})$ with $Y^2_{j,n} = (n-1)^H(X_H(\frac{i}{n-1}) - X_H(\frac{i-1}{n-1}))$. With Propositions 3.1 and 3.2 and Lemma 3.3, we have: $E_n(k) = (\frac{K}{n-1})^H E_k$ and

$$\begin{aligned}
\frac{S_n(k)}{E_n(k)} - 1 &= \frac{1}{E_k}\left(S_n(k)(\frac{n-1}{K})^{Hk} - E_k\right) \\
&= \frac{1}{(n-1)E_k}\left(\sum_{i=1}^{n-1}(|X_{i+1,n} - X_{i,n}|^k(\frac{n-1}{K})^{Hk} - E_k)\right) \\
&= \frac{1}{(n-1)E_k}\left(\sum_{i=1}^{n-1}(|Y^2_{i,n}|^k - E_k)\right) \\
&= \frac{A_n}{(n-1)E_k}T_n.
\end{aligned}$$

Hence, as $n \to \infty$,

$$(n-1)^\alpha\left(\frac{S_n(k)}{E_n(k)} - 1\right) \sim (n-1)^\alpha\frac{(n-1)^{1-\alpha}H(2H-1)}{(n-1)E_k}T_n = \frac{H(2H-1)}{E_k}T_n, \tag{4.8}$$

and, using Step 3 we get

$$n^\alpha \log(\frac{n}{K})(H_{n,K}(k) - H) \sim \frac{H(2H-1)}{E_k}T_n,$$

which combined with Theorem A and Lemma 4.1 yields (4.2).

Case B. The condition $\sum_{n\in\mathbb{N}} |c_0(n)|^l < \infty$ is equivalent to $H < 3/4$. Set $A^*_n = \sqrt{n}$ and $T^*_n = \frac{1}{A^*_n}\sum_{j=1}^{n-1}\mathbb{H}^k(Y^2_{j,n})$, following the same lines as above we get

$$\frac{S_n(k)}{E_n(k)} - 1 = \frac{A^*_n}{(n-1)E_k}T^*_n$$

and, using Step 3, we have as $n \to \infty$,

$$\sqrt{n}\log(\frac{n}{K})(H_{n,K}(k) - H) \sim \frac{1}{E_k}T^*_n,$$

which combined with Theorem B and Lemma 4.1 yields (4.1). □

Remark 4.3 - In each case, $0 < H < 3/4$ and $3/4 < H < 1$, we obtain asymptotic confidence intervals for our estimators via simulation experiments. The case $H = 3/4$ is not covered by the previous proposition. The following computation helps to illuminate the situation:

With the notation of Lemma 3.3, we denote by c_k the autocovariance function of the increments $(|Y_i|^k, 1 \leq i \leq n)$. Set $C_k = \frac{\pi}{2^k(\Gamma(\frac{k+1}{2}))^2}$ and $Z_n^k = \sqrt{n-1}(\frac{S_n(k)}{E_n(k)} - 1)$. We have

$$\begin{aligned} Var(Z_n^k) &= \frac{C_k}{(n-1)} . Var(\sum_{i=1}^{n-1} |Y_i|^k) \\ &= C_k . \{Var(|Y_1|^k) + \frac{2}{n-1} \sum_{i=1}^{n-2} (n-i-1) c_k(i)\} \\ &= C_k\{Var(|Y_1|^k) + 2\sum_{i=1}^{n-2} c_k(i) - \frac{2}{n-1} \sum_{i=1}^{n-2} i c_k(i)\} \\ &= C_k\{Var(|Y_1|^k) + \sum_{i=1}^{n-2} O(\frac{1}{i^{4(1-H)}}) - \frac{1}{n-1} \sum_{i=1}^{n-2} O(\frac{1}{i^{3-4H}})\}, \end{aligned}$$

where we have used Lemma 3.3.

By Lemma 3.4 we have that whenever $0 < H < 3/4$ there exists a $V_H(k) > 0$, such that $Var(Z_n^k) \rightarrow V_H(k)$, as $n \rightarrow \infty$ and whenever $3/4 < H < 1$ there exists $W_H(k) > 0$ such that, $Var(\frac{(n-1)^{2(1-H)}}{\sqrt{n-1}} Z_n^k) \rightarrow W_H(k)$ as $n \rightarrow \infty$. Finally, when $H = 3/4$, $Var(\sqrt{n-1}(\frac{S_n(k)}{E_n(k)} - 1)) \rightarrow \infty$ as $n \rightarrow \infty$.

- The computation above have the following by-product. If $H = 1/2$ then for each $k \geq 1$, we have $V_H(1) = \sqrt{\frac{\pi}{2}} \leq V_H(k)$. We thus conjecture that for $H \in (0,1)$, the best estimator might be given by $H_{n,K}(1)$.

With the notation of Proposition 3.2 and Lemma 3.5 we have

Proposition 4.4 *Set*

$$P_n = \frac{2}{\pi} \frac{S_n(2)}{(S_n(1))^2}.$$

Then we have almost surely

$$P_n \geq \frac{2}{\pi} \tag{4.9}$$

and we have

$$P_n \xrightarrow[n\to\infty]{P} 1, \tag{4.10}$$

$$\lim_{n\to\infty} P_{\nu_n} = 1 \quad a.s. \tag{4.11}$$

Proof. By Jensen's inequality we easily obtain (4.9).
We infer from Proposition 3.8, that, almost surely as $n \rightarrow \infty$

$$S_{\nu_n}(1) = (1+o(1))E_{\nu_n}(1) = \sqrt{\frac{2}{\pi}} \left(\frac{\nu_n - 1}{K}\right)^{-H} \quad \text{and} \quad S_{\nu_n}(2) = (1+o(1))E_{\nu_n}(2) = \left(\frac{\nu_n - 1}{K}\right)^{-2H},$$

which entail (4.11). Using the proof of Proposition 3.9, we know that $(\frac{S_n(1)}{E_n(1)}, \frac{S_n(2)}{E_n(2)}) \xrightarrow[n\to\infty]{P} (1,1)$ and by Slutsky's theorem, we finally obtain (4.10). □

Remark 4.5 The statistic P_n have the attractive properties to converge to 1 for correlated data and to be invariant by affine transformations of the data. Our simulations show that its value and the time series large fluctuations increase together, and thus, that it might be useful for practical applications. We suggest the name of "fractal variance" for the statistic P_n.

The following proposition establishes the asymptotic behavior of the fractal variance.

Proposition 4.6 *With the notation above,*
Whenever $H < \frac{7}{8}$ there exists a constant $V_H > 0$ such that

$$\sqrt{n}(P_n - 1) \xrightarrow{d} N(0, V_H) \tag{4.12}$$

and whenever $H > \frac{7}{8}$ we have for each $H < H' < 1$

$$n^{4(1-H')}(P_n - 1) \xrightarrow{d} 0. \tag{4.13}$$

Proof. **Step 1**. We will show that for each $H < H' < 1$

$$n^{\min\{1/2, 4(1-H')\}} \left(1 - \frac{S_n(1)}{E_n(1)}\right)^2 \xrightarrow{d} 0. \tag{4.14}$$

By Remark 4.3, we easily get

$$n^{\frac{\min\{1/2,4(1-H')\}}{2}} \left(1 - \frac{S_n(1)}{E_n(1)}\right) \xrightarrow{L^2} 0,$$

which obviously entails (4.14).

Step 2. We will show that for each $0 < H < 7/8$, there exists a constant $V_H > 0$ such that as $n \to \infty$

$$\sqrt{n}\left(\frac{S_n(2)}{E_n(2)} - 2\frac{S_n(1)}{E_n(1)} + 1\right) \xrightarrow{d} N(0, V_H) \tag{4.15}$$

and for each $7/8 < H < H' < 1$ we have

$$n^{4(1-H')}\left(\frac{S_n(2)}{E_n(2)} - 2\frac{S_n(1)}{E_n(1)} + 1\right) \xrightarrow{d} 0. \tag{4.16}$$

Set $\mathbb{H}^*(y) = E_1 y^2 - 2E_2|y| + E_1 E_2 = E_1(y^2 - E_2) - 2E_2(|y| - E_1)$, with as previously, $E_k = E(|Y|^k)$ $(k = 1, 2)$ where Y is a standard normal $N(0, 1)$ random variable. Our aim is to apply Theorems A and B, with $\mathbb{H} = \mathbb{H}^*$. In that view, we first compute $l = \min\{j \geq 1, a_j \neq 0\}$. Combining (4.5) and (4.6) we get

$$\begin{aligned}\mathbb{H}^*(y) &= E_1 E_2 H_2(y) - 2E_2\{\frac{1}{2}E_1 H_2 - \frac{1}{24}E_1 H_4(y)\} - 2E_2 \sum_{j=3}^{\infty} a_{2j} H_{2j}(y) \\ &= \frac{E_1 E_2}{12} H_4(y) - 2E_2 \sum_{j=3}^{\infty} a_{2j} H_{2j}(y) = \frac{1}{6\sqrt{2\pi}} H_4(y) - 2E_2 \sum_{j=3}^{\infty} a_{2j} H_{2j}(y).\end{aligned}$$

Thus we obtain

$$l = 4 = \min\{j \geq 1, a_j \neq 0\} \text{ and } a_l = \frac{1}{6\sqrt{2\pi}}.$$

We achieve our proof by considering two cases.

Case A. $\alpha = 2(1 - H) < 1/l$ which is equivalent to $H > 7/8$. In this case, following the same lines as in Step 3 of the proof of Proposition 4.2, we apply Theorem A, and we have $\frac{S_n(2)}{E_n(2)} - 2\frac{S_n(1)}{E_n(1)} + 1 = \frac{A_n}{(n-1)E_1 E_2} T_n$, which yields (4.16).

Case B. $\sum_{n\in\mathbb{N}} |c_0(n)|^l < \infty$ which is equivalent to $H < 7/8$. In this case, we apply Theorem B, and we have $\frac{S_n(2)}{E_n(2)} - 2\frac{S_n(1)}{E_n(1)} + 1 = \frac{A_n^*}{(n-1)E_1E_2}T_n^*$, which combined with the proof of Proposition 4.2 Step 3 Case B, yields (4.15).

Step 3. For each $0 < H < H' < 1$, we have

$$\begin{aligned} n^{\min\{1/2,4(1-H')\}}(P_n - 1) &= n^{\min\{1/2,4(1-H')\}}\frac{(E_n(1))^2}{(S_n(1))^2}\left(\frac{2}{\pi}.\frac{S_n(2)}{(E_n(1))^2} - \frac{(S_n(1))^2}{(E_n((1))^2}\right) \\ &= n^{\min\{1/2,4(1-H')\}}\frac{(E_n(1))^2}{(S_n(1))^2}\left(\frac{S_n(2)}{E_n(2)} - \frac{(S_n(1))^2}{(E_n((1))^2}\right) \\ &= n^{\min\{1/2,4(1-H')\}}\frac{(E_n(1))^2}{(S_n(1))^2}\left(\frac{S_n(2)}{E_n(2)} - 2\frac{S_n(1)}{E_n(1)} + 1 - (1 - \frac{S_n(1)}{E_n(1)})^2\right), \end{aligned}$$

which in combination with Proposition 3.8, Lemma 4.1, Step 1 and Step 2 completes the proof by taking suitable values for H'. □

Application 4.7 Under the general assumption that the data come from a discrete sample of a continuous process defined on $[0,1]$ with self-similar and stationary increments [*The increments of a random process* $\{Z(t), t \in \mathbb{R}\}$ *satisfy the self-similarity property with parameter* H ($H \geq 0$) *iff for each* $\tau \in \mathbb{R}$, $h > 0$ *and any* $t_0 \in \mathbb{R}$, $\{Z(t_0 + \tau) - Z(t_0)\} \stackrel{d}{=} \{h^{-H}[Z(t_0 + h\tau) - Z(t_0)\}$.], Proposition 4.6 gives a necessary condition satisfied by fractional Brownian motion and then allows to test the fit to the class of fractional Brownian motions. The choice of our general assumption was guided by experimental considerations.

5. Simulation Experiments

We have investigated the behaviour of the class of estimators of Proposition 3.2 in practical situations through a simulation study in which the parameter was allowed to vary through the entire parameter range. The random process was generated with the so called "Random Midpoints Displacement" method[1]. It is well known that the sampling process obtained is not exactly the one of a fractional Brownian motion. However, this method is widely used in the literature for generating fractional Brownian motion and assessing estimators of H (see e.g. Flandrin em et al (1991), Beran (1992)) mainly because it is a fast and simple one, that yields satisfactory results. Since this procedure is the most commonly used, we felt that we had to evaluate the performances of our estimator on samples generated with it.

The estimate $H_{n,K}(k)$ of H was computed for samples of size $n = 64$, $n = 256$ and $n = 1024$, and for $k = 1$, $k = 2$ and $k = 3$, where we always chose $K = 1$. The "fractal variance" (see Remark 4.5) was also evaluated.

We notice that using the "Random Midpoints Displacement" method, the optimal estimator is $H_{n,1}(2)$ contrary to our expectations in Remark 4.3.

$n = 64$, $K = 1$.

H	P_{64}	$H_{64,1}(1)$	$H_{64,1}(2)$	$H_{64,1}(3)$
0.1	1.01801	0.142922	0.130221	0.12464
0.2	0.885502	0.20654	0.214336	0.221275
0.3	0.711952	0.270068	0.285184	0.298637
0.4	0.887023	0.418	0.435041	0.443064
0.5	0.611117	0.56461	0.566267	0.563984
0.6	0.814732	0.612002	0.620918	0.624392
0.7	0.739843	0.71008	0.712436	0.716475
0.8	1.20175	0.817133	0.821131	0.823215
0.9	0.937915	0.786865	0.806371	0.819376
MSE	0.212646	0.0478634	0.0431569	0.0407099

[1] We use the "Algorithm Addition FM1D" taken from the book of Barnsley*et al* (1988) "The Sciences of Fractal images" (p.86).

$n = 256,\ K = 1.$

H	P_{256}	$H_{256,1}(1)$	$H_{256,1}(2)$	$H_{256,1}(3)$
0.1	0.937012	0.0886688	0.092278	0.0960624
0.2	0.881084	0.193322	0.193378	0.19288
0.3	0.86522	0.305391	0.30665	0.307923
0.4	1.16883	0.397214	0.393728	0.389638
0.5	0.92932	0.499755	0.50452	0.508319
0.6	0.99306	0.605898	0.607864	0.60901
0.7	0.786397	0.709935	0.708716	0.706519
0.8	0.962305	0.788773	0.793741	0.842416
0.9	0.842416	0.900075	0.899794	0.900077
MSE	0.125477	0.00716186	0.00653657	0.00693038

$n = 1024,\ K = 1.$

H	P_{1024}	$H_{1024,1}(1)$	$H_{1024,1}(2)$	$H_{1024,1}(3)$
0.1	1.05121	0.09940	0.10117	0.101887
0.2	1.05071	0.19496	0.197415	0.199382
0.3	1.00312	0.30166	0.303915	0.305331
0.4	1.02043	0.403471	0.4033	0.403226
0.5	0.942927	0.49987	0.499507	0.499336
0.6	0.966493	0.600982	0.599777	0.598698
0.7	0.983797	0.7031	0.703112	0.702643
0.8	1.08129	0.809038	0.809159	0.809388
0.9	0.835614	0.898758	0.900713	0.902896
MSE	0.0698367	0.00386406	0.00378067	0.00406082

We end our work based on simulation experiments by a compared study of numerical results taken from the paper of Flandrin *et al* (1991). In this paper, four methods for generating fractional Brownian motion were considered, including the "Random Midpoints Displacement" one. We only give here the figures concerning the "Random Midpoints Displacement" case, since it is the method that we used for computing our estimator. The reported estimates of H correspond to averages over five different realizations. In each case, different data lengths have been considered, except when analysis was clearly mismatched to the number of data points.

ESTIMATION METHODS	Number of samples	$H = 0.8$	$H = 0.5$	$H = 0.2$
Spectral analysis	64	0.42	0.45	0.18
	256	0.4	0.4	0.19
	1024	0.4	0.38	0.16
Higuchi	64	0.79	0.42	0.23
	256	0.81	0.47	0.26
	1024	0.8	0.51	0.25
Burlaga-Klein	64	0.78	0.38	0.02
	256	0.81	0.44	0.16
	1024	0.79	0.48	0.18
Maximum Likelihood Estimation	64	0.71	0.46	0.27
	256			
	1024			
Orthonormal wavelet analysis (Daubechies 6)	64			
	256			
	1024	0.59	0.38	0.17
Orthonormal wavelet analysis (Haar)	64			
	256			
	1024	0.75	0.39	0.09
Peltier - Lévy Véhel $H_{n,1}(1)$	64	0.83	0.51	0.2
	256	0.8	0.5	0.21
	1024	0.8	0.5	0.2

Notice that a comparison of performances of the different techniques involved by this analysis is made difficult by the fact there exists no unique fully satisfactory method to generating such processes. In the paper of Flandrin *et al* (1991), we notice that in most cases, Higuchi's method performs better than the others. We recall briefly this technique hereafter.

When considering a signal as a curve, self-similarity is revealed by the fact that its "length" (measured as the sum of absolute increments) for a given time interval varies as $L_H(k) = L_0 k^{-D}$, where $D = 2 - H$ is the fractal dimension of the curve and k is the number of data points. In the case of fractional Brownian motion, we have $E[L_H(k)] = L_0 k^{-D}$. Thus, if we compute the average $< L_H(k) >$ of all these lengths for time intervals containing k points, and if $< L_H(k) >$ is plotted against k on a doubly logarithmic scale, the data should fall on a straight line with a slope $-D$.

Thus, we see that Higuchi's technique based on the mean of absolute increments seems similar than our method. Hence, we think that in all cases our estimates should behave at least as efficiently as those of Higuchi's method.

The array above shows that $H_{n,1}(1)$ performs better than all other estimators.

CONCLUSION

In this work, we have proposed a new class of estimators whose main features are the following:

- the main statistical properties are known and in particular the rate of convergence is $O(1/\sqrt{n}\log n)$ in most cases.

- the practical implementation is straightforward, since it only involves the computation of the absolute increments of the process.

- tests on simulated data seem to show that the new estimator generally leads to better results than previously known techniques.

- at last, we have obtained an unexpected new fractal property of fractional Brownian motion in Theorem 3.7.

Let us mention that the specifity of our work is to have solved a statistical issue by means a fractal approach.

ACKNOWLEDGEMENT

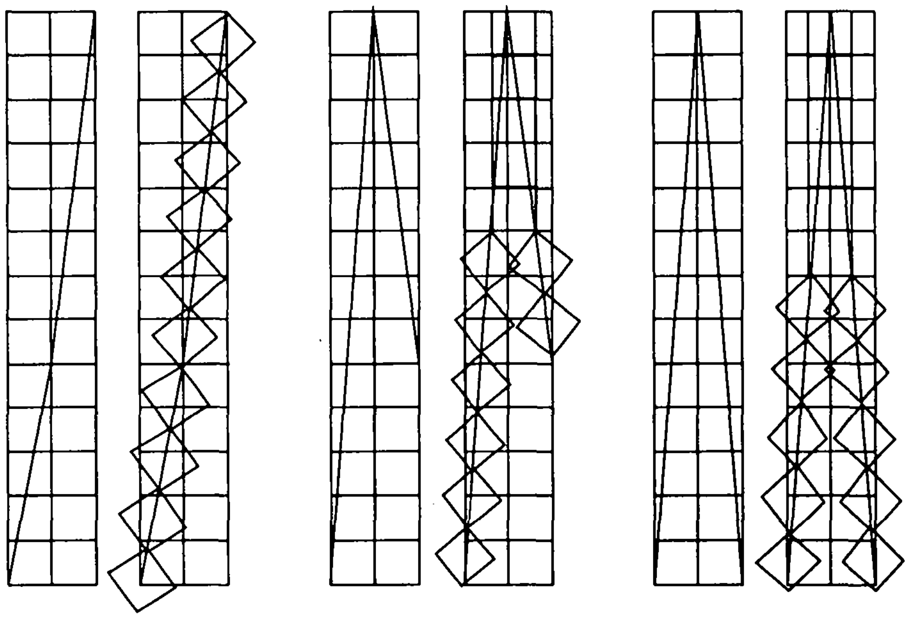
The authors would like to thank professor Deheuvels for its valuable comments on the earlier versions of this paper.

References

1. Beran J. Statistical methods for data with long-range dependence. *Statistical Science*, 7(4):404–427, 1992.
2. Barnsley M.F. Devaney R.L. Mandelbrot B.B. Peitgen H.-O. Saupe D. Voss R.F. . *The Science of Fractal Images*. Springer-Verlag, 1988.
3. Major P. Breuer P. Central limit theorems for non-linear functionals of Gaussian fields. *Journal of Multivariate Analysis*, 13:425–441, 1983.
4. Chung K.L. Williams R.J. *Introduction to Stochastic Integration*. 2nd. ed. Birkhäuser, 1990.
5. Dahlhaus R. Efficient parameter estimation for self-similar processes. *The Annals of Statistics*, 17(4):1749–1766, 1989.
6. Diaz J.P. *Processus non standard et queues de distribution*. Thèse de doctorat de mathématiques: Université de Paris 6, 1989.
7. Dobrushin R.L. and Major P. Non-central limit theorems for non-linear functionals of Gaussian fields. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 50:27–52, 1979.
8. El-Nouty. *Lois limites sur les processus fractals*. Thèse de doctorat de mathématiques: Université de Paris 6, 1994.
9. Falconer K. *Fractal geometry*. John Wiley and Sons., 1990.
10. Garreau D. Flandrin P., Gache N. Fractal dimension estimators for fractional Brownian motion. *IEEE Int. Conf. Acoust., Speech and Signal Processing, ICASSP-91, Toronto*, pages 3557–3560, 1991.
11. Fox R. and Taqqu M.S. Large-sample properties of parameter estimates for strongly dependent stationary Gaussian times series. *Ann. Statist.*, 14:517–532, 1986.
12. Leon J. Guyon X. Convergence en loi des h-variations d'un processus Gaussien stationnaire. *Ann. Inst. Henri Poincaré.*, 25:265–282, 1989. 3.
13. Stirzaker D.R. Grimmett G.R. *Probability and Random Processes*. Clarendon Press - Oxford, 1992.
14. Higuchi T. Approach to an irregular time series on the basis of the fractal theory. *Physica D*, 31:277–283, 1988.
15. Hurst H.E. Long-term storage capacity of reservoirs. *Trans. Amer. soc. Civil Engineers*, 116:770–808, 1951.
16. Hurst H.E. Methods of using long-term storage in reservoirs. *Proc. Institution of Civil Engineers*, I:519–577, 1955.
17. Loève M. *Probability theory II*. Springer Verlag, 1978.
18. Major P. Limit theorems for non-linear functionals of Gaussian sequences. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 57:129–158, 1981.
19. Mandelbrot B. *The fractal geometry of nature*. W.H. Freeman and Company. New York, 1983.
20. Mandelbrot B.B. and Van Ness J.W. Fractional Brownian motions, fractional Gaussian noises and applications. *SIAM Review 10*, 4:422–437, 1968.
21. Papoulis A. *Probability, Random Variables, and Stochastic Processes*. McDraw-Hill, 1991.
22. Taqqu M.S. Weak convergence to fractional Brownian motion and to Rosenblatt process. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 31:287–302, 1975.
23. Taqqu M.S. Convergence of integrated processes of arbitrary Hermite rank. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 50:53–83, 1979.

ANNEX A

In view of proving (3.15), we consider three cases as shown in Figure 1 below. These cases show the typical aspects of the graph of X_n on three consecutive time indices, together with the corresponding possible coverings by squares.



CASE 1 CASE 2 CASE 3

Figure 1

Elementary but lengthly geometric considerations allow us to say that (3.15) holds in all cases (case 1 is obviously true and case 3 represents the worst situation which corresponds to the biggest variation between $N^1_{\delta_n}(F_i)$ and N_i).

Unité de recherche INRIA Rocquencourt
Domaine de Voluceau - Rocquencourt - B.P. 105 - 78153 Le Chesnay Cedex (France)
Unité de recherche INRIA Lorraine - Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - B.P. 101 - 54602 Villers lès Nancy Cedex (France)
Unité de recherche INRIA Rennes - IRISA, Campus universitaire de Beaulieu 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes 46, avenue Félix Viallet - 38031 Grenoble Cedex 1 (France)
Unité de recherche INRIA Sophia Antipolis - 2004, route des Lucioles - B.P. 93 - 06902 Sophia Antipolis Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt - B.P. 105 - 78153 Le Chesnay Cedex (France)

ISSN 0249 - 6399

RR.2396